Onwubuya MN
Department of Statistics,
School of Applied Sciences and
Technology Delta State
Polytechnic, Otefe-Oghara,
Nigeria

Orighoyegha G
Department of Statistics,
School of Applied Sciences and
Technology Delta State
Polytechnic, Otefe-Oghara,
Nigeria

Analysis of macroeconomic predictive variables on gross domestic product using stepwise regression model

Onwubuya MN and Orighoyegha G

DOI: https://doi.org/10.22271/allresearch.2020.v6.i9f.7199

Abstract
Over the years, interest has been on estimating best regression models that minimizes error when several estimated models contain irrelevant independent variables. This study deals with analysis of macroeconomic predictive variables of gross domestic product. The data used for this study were obtained from a secondary source. A multiple regression model containing six independent variables of economic data were fitted. The variables included in the model are external debt, exchange rate, foreign direct investment, net export, consumption and debt payment. The data were analyzed through the use of stepwise regression with the help of statistical software (SPSS). The best model excluded both the debt service payment and the external for the single variable model and excludes only the debt service payment for the two variable models. Based on the analysis, consumption index and exchange rate influences the gross domestic product better; therefore, the use of stepwise regression is really essential in determining the predictive power of a regression model.

Keywords: Regression, predictive variable, Stepwise, gross domestic product, analysis

1. Introduction
Economic growth is that the priority of any credible government round the globe. Among the main macro-economic mandate of countries are to make sure a sustainable economic process, management of inflation and interest rate. The gross domestic product again may be a proxy for economic process because its dynamics significantly influences the general development of national economies. this is often the sum of gross value added by all resident producers within the economy alongside product of taxes and minus any subsidies not included within the value of the products. it's calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources (CBN 2014) [2]. Over the years, the appliance of multivariate analysis has gained acceptance in many discipline; however, decisions and predictions made could also be effective or misleading. Not all the variables proposed and wont to fit some models are all effective. That is, if the estimated model isn't carefully examined for adequacy, the results and inference are going to be ineffective in decision-making process. Then the method of stepwise regression method of eliminating all the irrelevant predictor variables from the model becomes paramount. Often, theory and knowledge give only general direction on which of a pool of candidate variables (including transformed variables) should be included within the regression model. the particular set of predictor variables utilized in the ultimate regression model must be determined by analysis of the information. Determination of this subset is named the variable selection problem. Finding this subset of independent variables involves two opposing objectives. First, let the regression model be as complete and realistic as possible. After knowing the connection between two variables, we could also be curious about estimating (predicting) the worth of 1 variable given the worth of another. The variable predicted on the idea of other variables is named the "dependent" or the 'explained' variable and therefore the other the 'independent' or the 'predicting' variable. The prediction is predicated on the average relationship derived statistically by multivariate analysis. The equation, linear or nonlinear is named the regression of y on x or the explaining equation. Thus, multivariate analysis reveals average relationship between two variables which makes possible estimation or prediction.

Corresponding Author:
Onwubuya MN
Department of Statistics,
School of Applied Sciences and
Technology Delta State
Polytechnic, Otefe-Oghara,
Nigeria

Consistent with Iwundu and Efezino (2015) [8], regression is that the measure of the typical relationship between two or more variables in terms of the first units of the information. Often, due to sample data, we wish to estimate the worth of a variable Y like a given value of a variable X. this will be accomplished by estimating the worth of Y from a least-squares curve that matches the sample data. The resulting curve is named a regression line of Y on X, since Y is estimated from X.

Jojo and Eric (2014) [10] on the opposite side stated that always times it's to estimate a regression of y on x that minimizes error and possess a high predictive power, by so doing we've to optimally estimate the model by scrutinizing the factor variables especially during a multi factor data analysis. Madu (2012) [12] noted that stepwise regression is an iterative technique that's wont to choose which predictor variables to incorporate during a regression model. Regression techniques are primarily utilized in order to make an equation which may be wont to predict values of dependent variables for all members of the population. It is often used as a way of explaining causal relationships between variables. Let every variable that's even remotely associated with the variable be included. Secondly, include a couple of variables as possible for every irrelevant variable decreases the precision of the estimated coefficients and predicted values. The presence of additional variables increases the complexity of knowledge collection and model maintenance. In data analysis, stepwise regression may be a method of fitting regression models during which the selection of predictive variables is administered by an automatic procedure. In each step, a variable is taken into account for addition to or subtraction from the set of explanatory variables supported some pre-specified criterion. Usually, this takes the shape of a sequence of F-tests or t-tests, but other techniques are possible, like adjusted R^2. The frequent practice of fitting the ultimate selected model followed by reporting estimates and confidence intervals without adjusting them to require the model building process under consideration has led to calls to prevent using stepwise model building altogether or to a minimum of confirm model uncertainty is correctly reflected.

Stepwise regression is one among this stuff like outlier detection and pie charts, which appear to be popular among some researchers but are considered by statisticians to be a touch of a joke. to deal with the difficulty more directly: the motivation behind stepwise regression is that you simply have tons of potential predictors but not enough data to estimate their coefficients in any meaningful way. the difficulty with stepwise regression is that, at any given step, the model is fit using unconstrained method of least squares. There are many various strategies for choosing variables for a regression model. If there are not any quite fifteen candidate variables, the All Possible Regressions procedure should be used since it'll always give nearly as good or better models than the stepping procedures available during this procedure. On the opposite hand, when there are quite fifteen candidate variables, the four search procedures contained during this procedure could also be of use. These search procedures will often find very different models. Outliers and collinearity can cause this. If there's little or no correlation among the candidate variables and no outlier problems, the four procedures should find an equivalent model. Hence, this study centers

on the utilization of stepwise regression in enhancing the predictive power of a regression model.

This paper is motivated by the relevance of the stepwise multivariate analysis as a notable approach in model fitting. Over the years, interest has been on estimating best regression models that minimizes error. However, several estimated models contain irrelevant independent variables. to assist the estimation of adequate and highly powered relationship model, there's got to include variable which will best describe the occurrence of the variable during this study, some macro-economic variables that describe economic process were wont to fit a multiple model and assessed using the stepwise regression approach. This research deals with the appliance of stepwise regression on the influence of macroeconomic variables on gross domestic products. to achieve this goal, the subsequent specific objectives are going to be useful to look at the appliance of stepwise regression in model variable selection, examine the effect of macro-economic variables on economic process and to scale back the fitted predictive model to best independent variables using the stepwise approach. This paper is dedicated to analyze the extent to which some factors affect GDP growth in developing countries as compared to developing countries.

There are some economic factors that emphasize all macroeconomic explanations of growth, possibly the foremost significant factor is that, so as to accumulate capital goods, the buyer goods will need to be foregone at the present to get more units of commodity within the future. So, a rise within the amount of capital goods or capital formation is termed as an investment; for economic process to occur, the extent of investment has got to be greater than the number of depreciation that's the quantity by which machines wear out or become outdated during the year. The greater the intensity of investment over depreciation, the larger the potential output of the economy within the future. Furthermore, economic processes are often investigated by adopting the concept of two components as in economic process a deviation or trade cycle and an economic trend component. The trend component or economic process is in charge of the long-term expansion and describes economic efficiency. The deviation component of economic process has got to have a zero mean within the end of the result.

In statistics, regression validation is that the process of deciding whether the numerical results quantifying hypothesized relationships between variables obtained from multivariate analysis are acceptable as descriptions of the information. The validation process can involve analyzing the goodness of fit of the regression, analyzing whether the regression residuals are randomly and checking whether the model's predictive performance deteriorates substantially when applied to data that weren't utilized in model estimation. how to check for errors in models created by step-wise regression isn't to believe the model's F-statistic, significance, or multiple R, but instead assess the model against a group of knowledge that wasn't wont to create the model. this is often done by building a model supported a sample of the info set available (e.g., 70%) and use the remaining 30% data set to assess the accuracy of the model. Accuracy is then often measured because the actual standard error (SE), or mean error between the anticipated value and therefore the actual value within the hold-out sample.

According Mikal (2012) [13] stepwise regression may be a method of fitting regression models during which the selection of predictive variables is administered by an automatic procedure. In each step, a variable is taken into account for addition to or subtraction from the set of explanatory variables supported some pre-specified criterion. Usually, this takes the shape of a sequence of F-tests or t-tests, but other techniques are possible, like adjusted R^2, Akaike information criterion, Bayesian information criterion, Mallows's C_p, or false discovery rate.

Stern and Coe (1984), opined that some approaches are: Forward selection, which involves starting with no variables within the model, testing the addition of every variable employing a chosen model fit criterion, add the varying quantities (if any) whose introduction resulted the simplest statistical significant enhancement of the fit. This process is repeated until none improves the model to an extent. Now, consider backward elimination, which involves starting with all candidate variables, testing the deletion of every variable employing a chosen model fit criterion, deleting the variable (if any) whose loss gives rise to the foremost statistical inconsequential worse growth of the model fit. This process is administered again until there's no more varying quantity which will be removed without a statistically significant loss of fit. In bidirectional elimination that's a mixture of the above, it deals with the testing at each step for variables to be included or excluded. Mundry and Fischer (1998) [6] noted that the forward selection may be a very attractive method because it's both tractable and provides an honest sequence of models.

- Start with a null model. The null model has no predictors, only one intercept (The mean over Y).
- Fit p simple rectilinear regression models, each with one among the variables in and therefore the intercept. So, basically you only search through all the single-variable models and therefore the best one (the one that leads to rock bottom residual sum of squares) you choose and fix it within the model.
- Now undergo the remaining p minus variables and determine which variable should be added to the present model to best improve the residual sum of squares.
- Continue until some stopping rule is satisfied, for instance when all remaining variables have a p-value above some threshold.

Unlike the forward stepwise selection Lawless, *et al.* (2018) [11] conducted a discourse using four independent variables with the approach of stepwise regression in optimizing the simplest model and that they stated: start with full method of least squares model that contains all p-predictors; then iteratively eliminate the less useful variable one-at-a-time. In backward elimination consistent with Cumt (2011) [3], it's the only of all variable selection steps which may be easily implemented without special software. Given a more complicated rank, backward elimination is often done manually as you're taking into consideration which varying quantities are susceptible to be eliminated. Here you begin with all the predictors within the model. Next, is deleting the predictor with highest p-value greater than α-critical value. Thus, refit the model and repeat the second step. Iteration is terminated when all p-values are but α-

critical value. The α-critical value also known a p-to-remove might not have got to be 5% and in testing prediction power thus, 15-20% limit will do better within the decision however, we should always note that the methods are designed directly for optimal forecast to be preferred. In forward selection, it's the reverse the backward method. It commences with no variables within the model and for each variable not within the expression, determine if their p-value are included to the model and choose the one with least p-value that's but α-critical value.

Mike and Charles (2010) [14] administered a study on the effect of engaging the approach of stepwise regression in model fitting and it had been revealed that it had been an inevitable means of evaluating regression models. Further, it had been opined that so as to be ready to do backward selection, it's necessary to be during a situation where there are more observations than varying quantities because you'll perform method of least squares regression when n is bigger than p. Supposing the worth of p is quite of n, we cannot fit a method of least squares model.

- Commence with every variable within the model.
- Eliminate the variable with the best p-value, that is, the unknown that have the smallest amount statistical significance.
- The new (p - 1) variable model is t, and therefore the variable with the most important p-value is removed.

This process continues until a stopping rule is achieved. for instance, you'll stop the iteration process when all remaining varying quantities have significant p-value that's defined by some significant threshold. Oladepo (2012) [15] stated that the procedures are indirectly linked to final goal of prediction or illustration thus, it's going to not actually assist solving the matter of interest. In using any predicator selection approach, it's crucial to recall that model selection can't be separated from the underlying objective under investigation. the choice of the variable tends to amplify the statistical significance of the variables that stay within the model. The variables that are dropped can still be correlated with the response. it might be wrong to mention these variables are unrelated to the response but it's just that they supply no additional explanatory effect beyond those variables already included within the model. The methodology of stepwise regression is employed in data processing, but controversial. Several points of criticism are made whereby tests are biased, since they're supported an equivalent data. The computed percentage points of the multiple regression coefficients by simulation showed that a final regression obtained by forward selection except for the F-procedure to be at 0.1%, was really only significant at 5%. In evaluating or estimating the degrees of freedom, the amount of independent varying quantities from the simplest fit selected is a smaller amount than the whole number of ultimate model varying quantities thereby causing the fit better than when adjusting the r2 value for the amount of degrees of freedom but it's pertinent to into consideration the amount of degrees of freedom that's utilized in the entire model.

2. Materials and Methods
This study is meant to function as explorative and inferential type. The paper described the factors which can influence gross domestic product of a nation and to form inference of

which of the factors is more significant. The populations encompass all the macroeconomic variables which can influence the turnover of the gross domestic product of an economy. Secondary source of knowledge was found useful because the data were sourced from the annual publication of the financial institution of Nigeria. the info used includes that of the determinant of GDP within the Nigerian economy obtained from the financial institution of Nigeria online database. They include determinant like: external debt, rate of exchange, consumption, foreign direct investment, net export and debt payment.

2.1 Method of Data Analysis

Stepwise regression is used to analyze the data obtained, stepwise essentially does multiple regressions a number of times each time removing the weakest correlated variable. At the end you are left with the variables that explain the distribution best. The only requirements are that the data is normally distributed (or rather, that the residuals are), and that there is no correlation between the independent variables.

The regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + e_{ij} \quad (1)$$

Where:
Y = Gross Domestic Product
X_1 = exchange rate
X_2 = external debt
X_3 = debt service payment
X_4 = foreign direct investment
X_5 = net export
X_6 = consumption
e_{ij} = Stochastic error term
β_0 = regression equation intercepts
$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5,$ and β_6 are regression coefficient

Suppose that there are p independent variables and n observations $(x_{11}, x_{12}, ..., x_{1p}, y_i)$, $i = 1, 2, ..., n$ and the model relating the independent variables to the dependent variable is specified in equation 1. Given the value of y denoted by y_i the system of the above equation becomes:

$$y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} + \varepsilon_1$$
$$y_1 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p} + \varepsilon_2$$
$$\vdots \quad \vdots \qquad \vdots \quad \vdots \qquad \vdots \quad \vdots$$
$$y_1 = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np} + \varepsilon_n$$

Expressing the above system in matrix form we obtain as follows;

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} \cdots & X_{1p} \\ 1 & X_{21} & X_{22} \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2)$$

The model is a system of n equations that can be expressed in matrix notation as defined:

$$y = X\beta + \varepsilon \quad (3)$$

We desire to obtain the vector of least squares estimator, $\hat{\beta}$, that minimizes the error sums of squares (SSE).

Using equation 3, make ε the subject of formular
$$\varepsilon = y - X\beta \quad (4)$$

The sum of squares of SSE

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \quad (5)$$

The least squares estimator $\hat{\beta}$ is the solution for β in the equations

$$\frac{\partial L}{\partial \beta} = 0$$
$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$
$$y'y - y'X\beta - X'\beta'y + X'\beta'X\beta \quad (6)$$

$$y'y - y'X\beta - y'X\beta + X'\beta'X\beta \quad (7)$$

$$\frac{\partial \varepsilon'\varepsilon}{\partial \beta} = 0 \rightarrow -y'X - y'X + 2X'X\beta$$
$$-2X'y + 2X'X\beta = 0$$
$$2X'X\beta = 2X'y$$
$$X'X\beta = X'y$$
$$\beta = \frac{X'y}{X'X}$$

$$\hat{\beta} = (X'X)^{-'}X'Y = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \quad (8)$$

2.2 Stepwise Regression

One common problem usually encountered by the users of multiple regression technique in analysis and other scientific studies is the selection of the right set of independent variables to be utilized in the model. The basic objective is how to find the "*best*" model. The model that do an acceptable and excellent job of predicting or explaining the dependent variable and at the same time uses the smallest possible set of independent variables.The stepwise regression is probably the most widely used search method not involving the computations of all possible regression equations. The first step in stepwise regression is to fit all simple regressions for each k independent variables. For each of such simple regression equation, test whether or not the slope is zero using the F test statistic.

$$F_k^* = \frac{MSR(X)_k}{MSE(X)_k} \quad (9)$$

The in-depth variable with the largest F^* value is test for significance using a predetermined F value. If it is significant, the independent variable is added, otherwise the process terminates with no independent variable considered sufficiently helped to enter the regression model.
Suppose X_4 is the independent variable that entered in step II, then all regressions with two independent variables having X_4 as one of the pair is calculated. The partial F test is used to test which of the recently included variable should be included in the model.

$$F_k^* = \frac{MSR(X_k/X_4)}{MSE(X_k/X_4)} \quad (10)$$

2.2.1 Forwards Selection Regression
The forward regression method of model selection puts variables into the equation, one at a time beginning with that

variable having the highest correlation (or R^2) with Y. For sake of argument, call this variable X_1. Next, it examines the remaining variables for the variable that when included with X_1 has the highest R^2. That predictor (with X_1) is inserted into the model. This procedure continues until adding the *best* remaining variable at that stage results in an insignificant increase in R^2 according to the partial F test.

2.2.2 Backward Elimination Regression
Backward elimination regression is the opposite of forward selection regression. It begins with all variables in the model and one by one removes them. It begins by finding the "*worst*" variable (the one that causes the smallest decrease in R^2 when removed from the complete model. If the decrease is insignificant, this variable is removed and the process continues.

2.3 Summary Procedure for Selecting the Best Independent Variables
The following steps are followed in selecting the best independent variables for a model:
- Fit a model having a single independent variable for all the available independent variables
- Compute the R^2 and the MSE for the models fitted above
- Locate the model having the highest R^2 and same time having the least MSE, that model then becomes the model with best independent variable.
- Repeat this for all the possible combination of a model having two independent variables, three, four, until the whole independent variables are combined so as to get the best independent variables for using two or three or four, etc…

The implication of isolating a given independent variable in the model on the performance of others implies that such independent variable cannot contribute to the predictive power of such model; hence it is then excluded from the model.

2.4 Hypothesis Testing
This study used the t-test of regression model parameter efficiency

The test statistic is defined as:

$$|t_0| = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{11}$$

Where $\hat{\sigma}^2 = MSE$, c_{jj} is the diagonal element of $(X'X)^{-'}$ corresponding to $\hat{\beta}_j$. Recall that $Se(\hat{\beta}_j)$ in equation 11 is the standard error of the regression coefficient $(\hat{\beta}_j)$. The degrees of freedom can be establish as $t_{\frac{\alpha}{2},n-p}$. Thus, H_0: β_j $=0$ is rejected if $|t_0| > t_{\frac{\alpha}{2},n-p}$.

Decision Rule: reject H_0 if p-value (sig) is less than level of significance, otherwise accept.

3. Data Presentation and Analysis
3.1 Data Presentation
Statistical data presentation is on exchange rate (Naira to a Dollar), external debt, GDP, debt service payments of the Nigerian economy, foreign direct investment, consumption and net export in billion dollars

Table 1: (Variable Data Presentation)

Years	Y	X₁	X₂	X₃	X₄	X₅	X₆
2000	46.385	102.1052	31.581	1.854	1.020	279	648.5
2001	44.138	111.9433	30.031	2.524	1.098	153	752.1
2002	59.116	120.9702	29.918	1.476	0.990	258	1101.2
2003	67.655	129.3565	34.136	1.631	1.892	111	1140.2
2004	87.845	133.5004	36.698	1.710	0.210	252	1602.5
2005	111.248	132.147	20.475	8.807	1.201	637	1972.3
2006	145.430	128.6516	3.964	6.710	0.987	619	2070.1
2007	166.451	125.8331	3.747	1.010	1.689	240	3072.1
2008	208.065	118.5669	4.042	0.412	1.225	817	7823.2
2009	169.481	148.8802	6.765	0.408	1.071	-141	8929.6
2010	369.062	150.298	7.206	0.292	1.504	4	11271.0
2011	411.744	153.8616	8.962	0.351	2.416	1068	11981.3
2012	462.979	157.4994	10.058	0.302	1.737	1779	13800.7
2013	521.803	157.3112	13.791	0.486	1.333	1470	14892.8
2014	871.944	171.7100	18.962	0.351	1.381	711	16215.9
2015	568.297	199.1120	21.058	0.502	1.213	510	18214.1
2016	481.070	198.482	12.991	0.686	0.673	-196	23840.39
2017	375.812	358.2	35.23	2.364	1.690	46.8	321551.4
2018	397.271	362.35	36.27	1.543	2.501	52.9	423226.7

Source: www.cbn.gov.ng/bulletton.pdf

3.2 Data Analysis
Using the Statistical Package (SPSS) software, the following procedure was followed to analyze the data
- Click on Analyze at top of the screen then
- Click on Regression then
- Click on Linear
- Highlight dependent variable by clicking on it and then
- Click on arrow > to transfer this name to the Dependent Box
- Highlight independent variables by clicking on it and then
- Click on arrow > to transfer this name to the Independent(s) Box
- Click on OK
- The results will appear in a Window. Scroll up using the slide bar on the right to the top of the output.

The SPSS output for the data analysis is presented below:

Table 2: (Analysis of variance for the multiple regression without stepwise)

ANOVAᵃ					
Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	739804.652	5	147960.930	9.620	.001ᵇ
Residual	169192.256	11	15381.114		
Total	908996.909	16			

a. Dependent variable: Gross Domestic Product (GDP)
b. Predictors: (Constant), Net export, Debt service payment, External debt, Exchange rate and Consumption

From above result, the model indicates adequate and fit for use and as over predictive power examination using Analysis of Variance (ANOVA)
In Table 3, results disagree with the result of table 2 where the entire predictive power is adequate, but in table 3, an individual observation of the parameter estimates indicates that only consumption was significant. Due to the result contradiction, there is need to apply the stepwise regression approach to verify which result is true.

Table 3: Coefficient estimates of the multiple regression without stepwise approach

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	**(Constant)**	-249.556	378.877		-0.659	0.524
	Exchange rate	2.252	3.562	0.262	0.632	0.540
	External debt	0.776	3.277	0.037	0.237	0.817
	Debt service Payment	-5.691	17.046	-0.057	-0.334	0.745
	Consumption	0.018	0.016	0.556	5.146	0.042
	Net export	0.114	0.063	0.259	1.813	0.097

a. Dependent Variable: GDP

3.3 Stepwise Analysis Result
The model summary statistics of the applied stepwise regression is given in the table below. The table displays the best possible model fit for the observed data, only two of the three explanatory variables were best for the fitting of the model

Table 4: Model Summary of stepwise method

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	Df 1	Df 2	Sig. F Change
1	0.632[a]	0.399	0.359	125136768080.67175	0.399	9.951	1	15	0.007
2	0.743[b]	0.551	0.487	111902147561.52567	0.152	4.758	1	14	0.047

a. Predictors: (Constant), consumption
b. Predictors: (Constant), consumption, exchange rate

Table 4 results indicated that the stepwise approach generated two models that best describe the process, one with one independent variable which the other with two independent variables and they both are significant.

Table 5: Analysis of variance on the models estimated by stepwise approach

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	155820.000	1	1558241900.000	9.951	0.007[b]
	Residual	2348880.000	15	156592107.000		
	Total	390712298.000	16			
2	Regression	21540299.000	2	10770149.000	8.601	0.004[c]
	Residual	17530926.000	14	12522090.000		
	Total	39071226.000	16			

a. Dependent Variable: GDP
b. Predictors: (Constant), consumption
c. Predictors: (Constant), consumption, exchange rate

The result of table 5 examines the overall predictive power of the new founded model from the stepwise approach and the result indicated that both models are adequate. It also explains that if only one independent variable is to be considered, the best will the consumption while if two variables are likely to be considered, the best should be the combination of consumption and exchange rate.

Table 6: The parameter estimates of the stepwise approach

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2882.065	164905412744.821		-1.748	0.101
	Consumption	356.740	1129198970.567	0.632	3.155	0.007
2	(Constant)	-108.300	168968861775.083		-0.641	0.532
	Consumption exchange rate	297.573	1045440781.703	0.527	2.842	0.013
		-5.495	2.519	-0.404	-2.181	0.047

a. Dependent Variable: GDP

The Result reported that if only one independent variable is to be considered, the best will the consumption while if two variables are likely to be considered, the best should be the combination of consumption and exchange rate. This result supports the findings in table 5 and table 3.

Table 7: Excluded variables from the model

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Exchange rate	-0.258[b]	-1.229	0.239	-0.312	0.882	1.134	0.882
	External debt	-0.404[b]	-2.181	0.147	-0.504	0.933	1.072	0.933
	Debt service payment	-0.280	-1.229	0.279	-0.512	0.582	1.334	0.782
	Net export	-0.304[b]	-2.181	0.947	-0.584	0.921	1.972	0.923
2	External debt	-0.255[c]	-1.378	0.192	-0.357	0.882	1.134	0.830
	Debt service payment	-0.121	-0.091	0.126	-0.215	0.612	1.213	0.612
	Net export	-0.812	-.1213	0.915	-0.229	0.281	2.001	0.281

a. Dependent Variable: GDP
b. Predictors in the Model: (Constant),
c. Predictors in the Model: (Constant), consumption, exchange rate

This summarizes the removed independent variables from the model due to their poor contribution to the predictive power of the model.

4. Backward Selection Outcome

4.1 Model for one explanatory variable
Only the consumption was found to be the only best explanatory variable when one explanatory variable is to be used to explain the GDP studied. Hence, the estimated regression model

$$\hat{Y} = -2882.065 + 356.740X_6 \tag{12}$$

4.2 Test of Hypothesis
Using the t-test to examine the above model, we have:
H_0: there is no significant relationship between consumption and GDP ($\beta_6 = 0$)
H_1: there is a significant relationship between consumption and GDP ($\beta_6 \neq 0$)
Level of significance ($\alpha = 0.05$)

$$\text{Test statistic} = |t| = \frac{\hat{\beta_6}}{\sqrt{\sigma^2 C_{ij}}} \rightarrow t_{\alpha, N-K}$$

Decision rule: reject H_0 if p-value (sig) is less than level of significance, otherwise accept.
Conclusion: since p-value (0.007) is less than $\alpha = 0.05$ we reject the null hypothesis and conclude that there is a significant relationship between consumption and GDP.

4.3 Model for two explanatory variables
Only the consumption and exchange rate and external debt was found to be the only best explanatory variables when two explanatory variables are to be used to explain the GDP studied. Hence, the estimated regression model is

$$\hat{Y} = -108.300 - 5.495X_1 + 297.573X_6 \tag{13}$$

4.3.1 Test of Hypothesis
Using the t-test to examine the above model, we have;
H_0: there is no significant relationship between exchange rate and GDP ($\beta_1 = 0$)
H_1: there is a significant relationship between exchange rate and GDP ($\beta_1 \neq 0$)
Level of significance ($\alpha = 0.05$)

$$\text{Test statistic} = |t| = \frac{\hat{\beta_1}}{\sqrt{\sigma^2 C_{ij}}} \rightarrow t_{\alpha, N-K}$$

Decision rule: reject H_0 if p-value (sig) is less than level of significance, otherwise accept.
Conclusion: since p-value (0.047) is less than $\alpha = 0.05$ we reject the null hypothesis and conclude that there is a significant relationship between consumption and GDP.

4.3.2 Test of Hypothesis
Using the t-test to examine the above model, we have;
H_0: there is no significant relationship between consumption and GDP ($\beta_6 = 0$)
H_1: there is a significant relationship between consumption and GDP ($\beta_6 \neq 0$)
Level of significance ($\alpha = 0.05$)

$$\text{Test statistic} = |t| = \frac{\hat{\beta_6}}{\sqrt{\sigma^2 C_{ij}}} \rightarrow t_{\alpha, N-K}$$

Decision rule: reject H_0 if p-value (sig) is less than level of significance, otherwise accept.
Conclusion: since p-value (0.013) is less than $\alpha = 0.05$ then reject the null hypothesis and conclude that there is a significant relationship between consumption and gross domestic product.

4.3.3 Model for three or more explanatory variables
There exists no possible model containing the entire independent variables. The variables dropped from the stepwise result, the p-value for all the t-test from the table output was more than the level of significance.

4.4 Discussion of Results
From table 2 result, the model indicates adequate and fit for use and as over predictive power examination using Analysis of Variance (ANOVA). Table 3 results above disagrees with the result of table 2 where the entire predictive power is adequate, but in table 4.3, an individual observation of the parameter estimates indicates that only consumption was significant. Due to the result contradiction, there is need to apply the stepwise regression approach to verify which result is true. Table 4 results indicated that the stepwise approach generated two models that best describe the process, one with one independent variable while the other with two independent variables and they both are significant. The result of table 5 examines the overall predictive power of the new founded model from the stepwise approach and the result indicated that both models are adequate. It also explains that if only one independent variable is to be considered, the best will the consumption

while if two variables are likely to be considered, the best should be the combination of consumption and exchange rate. The result of the analysis summarized here. The best model excluded the debt service payment, exchange rate, net export and the external for the single variable model and excludes only the debt service payment for the two variable models.

5. Summary and Conclusion

This study was set out to achieve the examination of the application of stepwise regression in model variable selection. Stepwise regression can be achieved either by trying out one independent variable at a time and including it in the regression model if it is statistically significant, or by including all potential independent variables in the model and eliminating those that are not statistically significant, or by a combination of both methods. To achieve this aim, data used for this study includes that of the determinant of GDP in the Nigerian economy. The data were analyzed through the use of SPSS package. The result indicated that the overall predictive power of the stepwise reduced model that if only one independent variable is to be considered, the best will the consumption while if two variables are likely to be considered, the best should be the combination of consumption and exchange rate. Based on the result of the study, it was concluded that in order to be certain of the predictive power of a regression model, there should be critical scrutiny of the model adequacy first. Notwithstanding, for this study, of all the six independent variables examined the model was reduced and concluded that consumption index and exchange rate influences the gross domestic product better. Based on the findings of this study it was recommended that, that researcher using multiple regression analysis should ensure the appropriate use of model adequacy as it helps in enhancing the predictive power of the fitted model. Also, further studies should examine more macroeconomic variables that may potentially affect the gross domestic product of a country and further studies should be carried out on examining the weakness of the stepwise regression and how to overcome them.

6. References

1. Bull KP, Bull DR: Model Selection and Multimodal Inference, 2nd Ed. Springer, Berlin, 2000.
2. Central Bank of Nigeria, 2014. www.cbn.gov.ng/bulletton.pdf
3. Cumt J. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 2nd Ed. Schaums Outline, 2001.
4. Dallal AJ. An Introduction to Generalized Linear Models, Chapman and Hall/Crc, Boca Raton, Florida, 2001.
5. Stern RD, Coe R. A Model Fitting Analysis of Daily Rainfall Data., Journal of The Royal Statistical Society. 1984; 147(1):1-34
6. Mundry R, Fischer J. Use of Statistical Programs for Nonparametric Tests of Small Samples Often Leads to Incorrect P-values: Examples from Animal Behavior, Animal Behaviour, 1998; 56(1):256-259.
7. Foster JJ. Extending Linear Models with R, Chapman and Hall/CRC, Boca Raton, Florida, 2001.
8. Iwundu PT, Efezino JT. The Use of an F-Statistic in Stepwise Regression Procedures, Technometrics. 2015; 14(2):182-194
9. Jack GP, Lazar MJ. Experimental Designs and Data Analysis for Biologists, 2nd Revised Edition, Cambridge University Press, Cambridge, United Kingdom, 2004.
10. Jojo JB. Eric KS. Model Selection in Ecology and Evolution, Data Mining Review of Economics and Statistics. 2014; 65(2):434-449
11. Lawless PM, Thompson WL, Kendall WR, Gould PF, Doherty Jr KP. Concerns Regarding a Call for Pluralism of Information Theory and Hypothesis Testing, Journal of Applied Ecology. 2008; 44(3):191-278
12. Madu A. Discovering Statistics Using SPSS. SAGE College Publishing, London, 2012.
13. Mikal MD. Arguments for Rejecting the Sequential Bonferroni in Ecological Studies, Oikos. 2002; 100(2):629-637
14. Mike FC, Charles D. Multivariate Analysis in Ecology and Systematics: Panacea or Pandora's Box, Annual Review of Ecology, Evolution, and Systematics. 2010; 21(1):233-246.
15. Oladepo S. A Farewell to Bonferroni: The Problems of Low Statistical Power and Publication Bias. Behavioral Ecology. 2012; 15(2):11-18.